学校编码：10384　　　　分类号 ______ 密级 ______
学号：17020051301618　　　　UDC ____________________

厦　门　大　学

硕　士　学　位　论　文

关于相依风险中多维联合测度的研究

On multivariate measures of association for dependent risks

李 玉 水

指导教师姓名：张 志 强 副教授

专 业 名 称：概率论与数理统计

论文提交日期：2008 年 5 月

论文答辩日期：2008 年　月

学位授予日期：2008 年　月

答辩委员会主席：______________

评　　阅　　人：______________

2008 年 5 月

厦门大学学位论文原创性声明

兹呈交的学位论文，是本人在导师指导下独立完成的研究成果．本人在论文写作中参考的其它个人或集体的研究成果，均在文中以明确方式标明．本人依法享有和承担由此论文而产生的权利和责任．

声明人（签名）：

年　　月　　日

厦门大学学位论文著作权使用声明

本人完全了解厦门大学有关保留、使用学位论文的规定.厦门大学有权保留并向国家主管部门或其指定机构送交论文的纸质版和电子版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆被查阅，有权将学位论文的内容编入有关数据库进行检索，有权将学位论文的标题和摘要汇编出版.保密的学位论文在解密后适用本规定.

本学位论文属于

1、保密（　　），在　　　年解密后适用本授权书.

2、不保密（　　）.

（请在以上相应括号内打"✓"）

作者签名：　　　　　　　　日期：　　　年　月　日

导师签名：　　　　　　　　日期：　　　年　月　日

目　录

Contents

摘　要

在概率论与统计学中, 为了简化要研究的问题, 往往忽略了随机变量之间复杂的相依关系 (尤其是多维随机变量之间常常存在更为复杂的相依关系), 假定它们之间相互独立, 但这种忽略所得到的结论不符合实际情况. 鉴于实际中存在的这种问题, 对随机变量之间相依性的研究就显得更为重要. 本文就是从随机变量之间的相依关系入手进行研究, 在前人研究的基础上对测度 Kendall's τ, Spearman's ρ, Blomqvist's β, Gini's γ 在多维和谐概念的基础上进行多维上的推广以及讨论了不同维度之间的关系. 具体内容如下：第一章, 介绍了本文的写作背景并简要介绍本文的主要研究工作. 第二章, 给出了必要的基础知识. 第三章主要讨论和谐性在多维下的定义及基于多维和谐性下的联合测度的推广, 并且分析推广后多维联合测度的性质. 第四章中给出一些具体的应用及有趣的结论.

关键词: 联合测度；和谐性；连接函数

Abstract

In light of complexity of correlations between random variables(especially multi-dimensional random variables) in probability theory and statistics, it is invariably postulated that random variables are independent from one another. Nevertheless, in practice absolute independence is hard to be satisfied. Moreover, the conclusion we reach, after ignoring dependence of random variables, can not conform to the fact. The paper improves the measurement which depict the concordance from the dependence among the random variables, on the basis of the previous studies, we give the extensions to the multivariate measures (i.e. Kendall's τ, Spearman's ρ, Blomqvist's β, Gini's γ) of association that base on the concept of multi-concordance between two random vectors and discuss the relationship between different dimension's measures of association. The structure of the paper is as follows.

In chapter 1, we present the background of the paper and simply introduce the main research; In chapter 2, we introduce basic definitions and properties used in the sequel; In chapter 3, we give the definition of multi-concordance and provide extensions of Kendall's τ, Spearman's ρ, Blomqvist's β, Gini's γ based on multi-concordance. In chapter 4, we study applications of multi-dimensional measures of association and some examples.

Key words: Measures of association; Concordance; Copulas.

第一章 引 言

在概率论和统计学中, 为了便利于研究和具体的计算, 总是假定随机变量之间相互独立. 实际上它们之间却往往存在着复杂的关系, 我们把这种复杂的关系总称为随机变量之间的相依关系. 在文献 [1] 中给出了相依的概念. 由于相依关系会对我们的研究及其计算结果造成影响, 鉴于它的重要性, 所以对它的研究一直以来是个热点问题. 有很多研究从不同角度对随机变量相依性度量指标进行刻画, 用度量值来衡量随机变量之间的相依性. 常用的有相关系数, 协方差等, 但是其中有些度量指标刻画的过于粗糙. 例如相关系数只是用一个数值来刻画总体的线性相关指标, 若此指标为零, 说明随机变量之间没有线性关系, 对于是否存在其它的关系就没有刻画.

随机变量相依性的研究较长时间没有大的进展, 其中一个重要的原因是没有好的数学工具可以借助. 随着概率积分变换, Copula 的出现, 使得对随机变量的相依性研究又迎来一个新的春天, 相依性研究向前迈进了一大步. 本文主要是借助 Copula 对随机变量的相依关系进行研究.

利用 Copula 对随机变量相依性的最早文献是 Scliweiezr 和 Wofll 的文章, 他们讨论了两个随机变量之间的相依测度在随机变量严格单调变换下的 Copula 不变性. 有很多利用 Copula 研究随机变量相依性的文献. 现在对相依关系的研究多在建立新的相依性度量指标, 利用这些指标从不同的侧面对随机变量之间的相依关系进行刻画. 由这些数据来判定随机变量之间呈现哪种相依关系. 从二十世纪五十年代至今, 出现了下列的度量指标, 主要是两个随机变量之间的和谐测度 Kendall's τ, Spearman's ρ, Blomqvist's β, Gini's γ 等. 但是这些相依性度量指标也有很多不完善的地方, 使得对随机变量相依性的较深层次的研究依然处于初级阶段. 多维随机变量之间常常存在着复杂的相依关系, 但大多数方法刻画其相依的方法都是用了两两之间相依测度构成的矩阵, 而忽视从整体上用一个相依值去刻画其相依的大小. 因此有必要对上述的四个和谐测度进行推广到 n 个随机变量的情景. 本文就是在此基础之上, 对原有的相依性度量指标进行推广, 更深一步的挖掘随机变量之间的相

依性，对解决实际问题提供帮助.

本文主要工作是：对测度 Kendall's τ, Spearman's ρ, Blomqvist's β, Gini's γ 在多维和谐概念的基础上进行多维上的推广以及讨论了不同维度之间的关系. 在第二章节中主要介绍了本文所需的 Copula 和和谐性及其测度的一些基础知识. 第三章主要讨论和谐性在多维下的定义及基于多维和谐性下的和谐测度的推广，并且分析推广后的多维联合测度性质. 第四章中给出一些具体的应用及有趣的结论.

第二章 预备知识

§2.1 连接函数（ Copula ）

随机变量之间的相依关系主要蕴涵在联合分布函数中，由边缘分布无法唯一确定联合分布, 因此若想借助于边缘分布来讨论相依性很难. 联合分布的每个自变量的取值为整个实数轴, 使所要研究的问题更加显得复杂. 如何通过边缘分布函数来研究相依关系呢？并且把每个自变量的取值变为某个特定区间上的取值？“ Copula ”就可以解决这些问题，它可以把联合分布函数用其一维的边缘分布函数表示，把联合分布的定义域，即整个平面转化为一个边长为 1 的正方形区域. 对随机变量的相依性的研究就可以转化为对边际分布函数之间关系的研究，利用 Copula 对相依性的研究就成为数学家关注的对象.

“ Copula ”是一个拉丁语名词. 意为“连接，系，捆绑”等含义. 它实质上就是连接联合分布函数与边际分布函数的函数. 现在国内的很多文献把它称为连接函数. 在本文中仍采用原文“ Copula ”来表示.

对这个问题研究最早在 1940 和 1941 年， Hoedffnig 发现“标准化函数”及其一些结论. 1951 年 Fréchet 也取得了很多和 Hoedffnig 相同的结论，还发现了“ Fréchet-Hoeffding 界”等重要的性质. 在 1959 年 Sklar 第一次把 Copula 这个定义引入到数学中来，并且证明了“ Copula ”的存在性. 二十世纪七十年代，Kimeldorf 和 Smapson, Galmabor 和 Deheuvels[1978] 也找到了 Copula 函数，只是名字不同. Sklar 和 Schwezier 合作研究概率度量空间也得到很多关于 Copula 的重要结论. 通过在二十世纪九十年代这十年四次著名的国际数学会议，“ Copula ”在统计与概率的应用中得到发展. Nelesn 在 1999 年把 Copula 的定义及其重要的结论做了系统总结，从而使 Copula 这个概念得以进一步的推广. 现在 Copula 还用在随机过程的研究中. 更多领域正在引入 Copula 并应用它，而它也越来越焕发出勃勃生机.

§2.1.1 Copula **函数定义、性质**

Copula 被 Sklar(1959) 首次引入, 用于统计学, 是表示将一元分布函数 "连接" 起来形成多元分布函数. 为了给出它的定义和基本性质，首先给出几个预备概念.

$\overline{R}$ 表示扩展后的实轴 $[-\infty, +\infty]$, $S_1, S_2, \cdots, S_n$ 是 $\overline{R}$ 的非空子集， H 是定义在 $DomH = S_1 \times S_2 \times \cdots \times S_n$ 上的实函数，向量 $\mathbf{a} = (a_1, a_2, \cdots, a_n)$,$\mathbf{b} = (b_1, b_2, \cdots, b_n)$. 并且对任意的 k 有 $a_k \leq b_k$. 令 $B = [a_1, b_1] \times \cdots \times [a_n, b_n]$,B 的所有顶点都在 $DomH$ 内. $\mathbf{c} = (c_1, c_2, \cdots, c_n)$ 表示 B 的顶点，定义

$$V_H(B) = \sum sgn(\mathbf{c})H(\mathbf{c}),$$

其中

$$sgn(\mathbf{c}) = \begin{cases} 1 & \text{若有偶数个 } c_k = a_k \\ -1 & \text{若有奇数个 } c_k = a_k. \end{cases}$$

(1) 如果 $V_H(B) \geq 0$ 成立，其中 B 的所有顶点都落在 DomH 内，则称 H 是 n 元增函数.

(2)H 是定义在 $DomH = S_1 \times S_2 \times \cdots \times S_n$ 上的实函数，假设每个 S_k 都有一个最小的元素 a_k, 对于任意的向量 $\mathbf{t} = (t_1, t_2, \cdots, t_n) \in DomH$, 若存在一个 $t_k = a_k$ 便有 $H(\mathbf{t}) = 0$ 成立，那么则称 H 是基础的.

(3) 若每个 S_k 是非空的和都有一个最大元素 b_k, 则函数 H 有边缘函数而且其一维的边缘函数 H_k 满足 $DomH_k = S_k$ 以及 $H_k(x) = H(b_1, \ldots, b_{k-1}, x, b_{k+1}, \ldots, b_n)$ 对任意的 $x \in S_k$ 都成立. 类似地，我们也可以定义更高维度的边缘函数。

引理 **2.1** 若 $S_1, S_2, \cdots, S_n$ 是 $\overline{R}$ 的非空子集， H 是定义在 $DomH = S_1 \times S_2 \times \cdots \times S_n$ 上的实函数且是基础的 n 元增函数. 则 H 是关于每个变量是递增的, 即对任意的 $(t_1, \ldots, t_{k-1}, x, t_{k+1}, \ldots, t_n), (t_1, \ldots, t_{k-1}, y, t_{k+1}, \ldots, t_n) \in DomH$ 及 $x \leq y$, 都有 $H(t_1, \ldots, t_{k-1}, x, t_{k+1}, \ldots, t_n) \leq H(t_1, \ldots, t_{k-1}, y, t_{k+1}, \ldots, t_n)$.

引理 **2.2** 若 $S_1, S_2, \cdots, S_n$ 是 $\overline{R}$ 的非空子集， H 是定义在 $DomH = S_1 \times S_2 \times \cdots \times S_n$ 上的实函数且是基础的 n 元增函数. 则对于任意的 $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} =$

$(y_1,\ldots,y_n)\in S_1\times S_2\times\cdots\times S_n$, 都有

$$\mid H(\mathbf{x})-H(\mathbf{y})\mid\leq\sum_{k=1}^{n}\mid H_k(x_k)-H_k(y_k)\mid.$$

其证明过程可查阅 $Schweizer, Sklar(1983)$ 相关文献.

定义 **2.1** 若 H 是定义在 $DomH=S_1\times S_2\times\cdots\times S_n$ 上基础的 n 元增函数. 且 $H(\infty,\ldots,\infty)=1$. 则称 H 是 n 维的分布函数.

由引理 2.1 可得：任一 n 维的分布函数其一维的边缘函数也是分布函数，我们在这把它们记为 $F_1,\ldots,F_n$.

定义 **2.2** 一个 n 维的 Copula 是定义在 $[0,1]^n$ 上的实函数，以符号 C 表示，并满足以下二个条件：

(a). C 是基础的、 n 元增函数;

(b). C 的所有边缘分布 C_i 满足： $C_i(u)=C(1,\cdots,1,u,1,\cdots,1)=u$, 其中 $u\in[0,1]$, $i=1,2,\ldots,n$.

对于任一个 n-copula $C, n\geq 3$, 其每一 k 维的边缘分布是 k-copula. 或者说一个 n-copula C 是 $[0,1]^n\to[0,1]$ 的函数且满足以下两个性质：

1. 对任 $\mathbf{u}\in[0,1]^n$， $C(\mathbf{u})=0$ 当 $\mathbf{u}$ 至少有一个分量为 0 时， $C(\mathbf{u})=u_k$ 当 $\mathbf{u}$ 的所有的分量除 u_k 以外都为 1 时.

2. 对于任意的 $\mathbf{a}$, $\mathbf{b}\in[0,1]^n$, 若满足 $a_i\leq b_i$, $\forall$ i, 有 $V_C([\mathbf{a},\mathbf{b}])\geq 0$.

因为 n-copula 是 $[0,1]^n$ 上的分布函数，所以其 $[0,1]^n$ 在概率测度可由以下式子确定：

$$V_C([0,u_1]\times\cdots\times[0,u_n])=C(u_1,\ldots,u_n).$$

这个概率测度就记为 V_C.

§2.1.1 Sklar **定理**

以下定理就是著名的 Sklar 定理，它是关于 Copula 理论中最重要的结论，也是 Copula 在各个方面应用中最有意义的。

定理 2.1 若 H 是一个 n 维联合分布函数，其边缘分布函数为 $F_1,\ldots,F_n$. 那么一定存在一个 n-copula 函数 C，使得：对所有的 $\mathbf{x}\in\bar{R}^n$, 有

$$H(x_1,\ldots x_n)=C(F_1(x_1),\ldots,F_n(x_n)). \tag{2.1}$$

如果边缘分布 $F_1,\ldots,F_n$ 是连续的, 那么 Copula 形式是唯一的. 否则 C 在 $RanF_1$, $\ldots,RanF_n$ 上就不能唯一确定. 反过来，如果 C 是 Copula 函数， $F_1,\ldots,F_n$ 边缘分布函数，那么 H 就是边缘分布为 $F_1,\ldots,F_n$ 的联合分布函数. Sklar 定理的证明可查阅 Sklar(1996) 文献.

若 F 单变量的分布函数，我们定义 F 的广义逆函数为 $F^{-1}(t)=inf\{x\in R\mid F(x)\geq t\}\ \ \forall t\in[0,1]$, 并规定 $inf\emptyset=-\infty$.

引理 2.3 若 H 是一个 n 维连续的联合分布函数, 其边缘分布函数为 $F_1,\ldots,F_n$ 及 Copula C (这时 C 满足式 (2.1)). 那么对 $\forall\mathbf{u}\in[0,1]^n$, 有

$$C(u_1,\ldots,u_n)=H(F_1^{-1}(u_1),\ldots,F_n^{-1}(u_n)).$$

若没有连续的条件, 则结论就不一定成立. 可查阅 Nelsen (1999) 或 Marshall (1996) 文献.

例子 2.1 若 Φ 表示标准的单变量正态分布函数， Φ_R^n 表示标准的 n 变量正态分布函数，而且它的线性相关矩阵为 R. 则

$$C(u_1,\ldots,u_n)=\Phi_R^n(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_n)).$$

是高斯或正态 n-copula.

§2.1.3 Copula 函数的 “ Fréchet-Hoeffding 界”

首先定义 $[0,1]^n$ 上的三个函数 M^n, Π^n 和 W^n 如下：

$$M^n(\mathbf{u})=min(u_1,\ldots,u_n),\Pi^n(\mathbf{u})=u_1\cdots u_n,W^n(\mathbf{u})=max(u_1+\cdots+u_n-n+1,0).$$

其中函数 M^n 和 Π^n 对 $\forall n\geq 2$ 都是 n-copulas，而函数 W^n 对 $\forall n\geq 3$ 则不是一个 Copula, 我们来看下面这个例子.

例子 2.2 考虑 n- 体 $[1/2,1]^n \subset [0,1]^n$.

$$\begin{aligned} V_{W^n}([1/2,1]^n) &= max(1+\cdots+1-n+1,0) \\ &\quad -nmax(1/2+1+\cdots+1-n+1,0) \\ &\quad +\binom{n}{2}max(1/2+1/2+1+\cdots+1-n+1,0)\cdots \\ &\quad +max(1/2+\cdots+1/2-n+1,0) \\ &= 1-n/2+0+\cdots+0. \end{aligned}$$

因此 W^n 不是一个 copula 当 $n \geq 3$ 时.

下面的定理是 n 维 Copula 函数的 Fréchet-Hoeffding 界的不等式.

定理 2.2 如果 C 是任一个 n-copula, 那么对于 $[0,1]^n$ 中的任一向量 $\mathbf{u}$, 都有

$$W^n(\mathbf{u}) \leq C(\mathbf{u}) \leq M^n(\mathbf{u}).$$

更多的细节，包括其几何意义，可查阅 Mikusinski, Sherwood, Taylor(1992) 等文献. 尽管 FréchetHoeffding 的下界 W^n 不是一个 copula 当 $n \geq 3$ 时，但它却是在以下意义下最好的下界.

定理 2.3 对于 $\forall \mathbf{u} \in [0,1]^n$ 且 $n \geq 3$, 存在一个 n-copula C(其依赖于 $\mathbf{u}$), 使得

$$C(\mathbf{u}) = W^n(\mathbf{u}).$$

其证明过程可见 Nelsen(1999)p.42.

我们记 $\bar{C}$ 为 n 个随机变量的联合生存函数，且这 n 个随机变量的联合分布函数为 C, 即若 $(U_1,\ldots,U_n)^T$ 的分布函数是 C, 则有 $\bar{C}((u_1,\ldots,u_n) = P\{U_1 > u_1,\ldots,U_n > u_n\}$.

定义 2.4 设 C_1 和 C_2 都是 n 维 Copula 函数，如果对于 $[0,1]^n$ 中任意的 $\mathbf{u}$ 都有

$$C_1(\mathbf{u}) \leq C_2(\mathbf{u}), \quad \bar{C}_1(\mathbf{u}) \leq \bar{C}_2(\mathbf{u}).$$

我们称 C_1 小于 C_2(或者说 C_2 大于 C_1)，我们记做 $C_1 \prec C_2$(或者 $C_1 \succ C_2$). 注意对于双变量有如下结论:

$$\begin{aligned}\bar{C}_1(u_1,u_2) \le \bar{C}_2(u_1,u_2) &\Leftrightarrow 1-u_1-u_2+C_1(u_1,u_2) \le 1-u_1-u_2+C_2(u_1,u_2)\\ &\Leftrightarrow C_1(u_1,u_2) \le C_2(u_1,u_2).\end{aligned}$$

因此, Fréchet-Hoeffding 下界 W^n 小于任何一个 n 维 Copula 函数, 而 Fréchet-Hoeffding 上界又大于任何一个 n 维 Copula 函数. 对于含参数的 Copula 函数, 比如单参数连接函数集 C_θ, 如果对于任何的 $\theta_1 \le \theta_2$, 有:

$$C_{\theta_1} \prec C_{\theta_2}.$$

那么, 我们称 Copula 函数集 C_θ 是正序的; 相反地, 如果 $\theta_1 \le \theta_2$, 有:

$$C_{\theta_1} \succ C_{\theta_2}.$$

那么我们称 Copula 函数集 C_θ 是逆序的.

§2.1.4 Copula **函数和随机变量**

因为 $X_1,\ldots,X_n$ 相互独立当且仅当对于 $\forall x_1,\ldots,x_n \in \bar{R}$, 有 $H(x_1,\ldots,x_n) = F_1(x_1)\cdots F_n(x_n)$, 则以下结论可从定理 2.1 的得到.

定理 **2.4** 若 $(X_1,\ldots,X_n)^T$ 是由连续的随机变量构成的 n 维向量, 其 copula 函数为 C, 则 $X_1,\ldots,X_n$ 相互独立当且仅当 $C=\Pi^n$.

Copula 函数的一个重要性质是对于随机变量的严格单调变换具有不变性或者以简单形式改变. 注意到, 如果随机变量 X 的分布函数是连续的, 并且函数 α 在 $RanX$ 上是严格单调函数, 那么新的随机变量的 $\alpha(X)$ 的分布函数也是连续的.

定理 **2.5** $(X_1,\ldots,X_n)^T$ 是由连续的随机变量构成的 n 维向量, 其 Copula 函数为 C. 若 $\alpha_1,\ldots,\alpha_n$ 分别在 $RanX_1,\ldots,RanX_n$ 是严格的单调递增函数, 则 $(\alpha_1(X_1),\ldots,\alpha_n(X_n))^T$ 的 Copula 函数也是 C.

证明: 设 $F_1,\ldots,F_n$ 分别表示 $X_1,\ldots,X_n$ 的分布函数， $G_1,\ldots,G_n$ 分别表示函数 $\alpha_1(X_1),\ldots,\alpha_n(X_n)$ 的分布函数. C 是随机向量 $(X_1,\ldots,X_n)^T$ 的 Copula 函数， C_α 是 $(\alpha_1(X_1),\ldots,\alpha_n(X_n))^T$ 的 Copula 函数，由于每个 α_k 都是严格单调增的，对 $\bar{R}$ 中的任意 x, 有 $G_k(x)=P\{\alpha_k(X_k)<x\}=P\{X_k\le\alpha_k^{-1}(x)\}=F_k(\alpha_k^{-1}(x))$, 所以

$$\begin{aligned}
C_\alpha(G_1(x_1),\ldots,G_n(x_n)) &= P\{\alpha_1(X_1)<x_1,\ldots,\alpha_n(X_n)<x_n\}\\
&= P\{X_1\le\alpha_1^{-1}(x_1),\ldots,X_n\le\alpha_n^{-1}(x_n)\}\\
&= C(F_1(\alpha_1^{-1}(x_1)),\ldots,F_n(\alpha_n^{-1}(x_n)))\\
&= C(G_1(x_1),\ldots,G_n(x_n)).
\end{aligned}$$

由于 $X_1,\ldots,X_n$ 是连续的， $RanG_1=\cdots=RanG_n=[0,1]$, 那么我们得出在 $[0,1]^n$ 中，有 $C_\alpha=C$.

从 Sklar 定理，我们得出 Copula 函数 C 把 n 维分布函数分成了边缘分布和一个表示相关结构的函数. 以下的这个定理表明，也同样存在一函数 $\hat{C}$ ，将 n 维生存函数分解成单一的边缘分布和一个表示相关结构的函数。进一步可以证明出这个函数也是 Copula 及这个生存 Copula 函数能表示成 C 和它的 k 维边缘分布的形式.

定理 **2.6** 设 $X_1,\ldots,X_n$ 是连续的随机变量，有 Copula 函数 $C_{X_1,\ldots,X_n}$. 函数 $\alpha_1,\ldots,\alpha_n$ 分别在 $RanX_1,\ldots,RanX_n$ 上是严格单调函数，且设 $\alpha_1(X_1),\ldots,\alpha_n(X_n)$ 有 Copula 函数 $C_{\alpha_1(X_1),\ldots,\alpha_n(X_n)}$. 更进一步，我们假定对于某一个 k ， α_k 是一个严格递减函数，此处 $1\le k\le n$ ，不失一般性，假设 $k=1$ ，那么

$$\begin{aligned}
C_{\alpha_1(X_1),\ldots,\alpha_n(X_n)}(u_1,u_2,\ldots,u_n) &= C_{\alpha_2(X_2),\ldots,\alpha_n(X_n)}(u_2,\ldots,u_n)\\
&\quad -C_{X_1,\alpha_2(X_2),\ldots,\alpha_n(X_n)}(1-u_1,u_2,\ldots,u_n).
\end{aligned}$$